METHOD AND APPARATUS FOR READING REPORTER LABELED BEADS

Related Applications

This application is based on prior copending provisional application Serial No. 60/240,125, filed on October 12, 2000, and prior copending provisional application Serial No. 60/242,734, filed on October 23, 2000, the benefits of the filing dates of which are hereby claimed under 35 U.S.C. § 119(e). This application is further based on prior copending conventional application No. 09/939,292, filed on August 24, 2001, which is based on a prior copending provisional application Serial No. 60/228,076, filed on August 25, 2000, the benefits of the filing dates of which are hereby claimed under 35 U.S.C. § 119(e). and under 35 U.S.C. § 120.

Field of the Invention

The present invention generally relates to a method and apparatus employed to image an encoded reporter labeled bead, and more specifically, to a method and apparatus that enable encoded reporter labeled beads to be imaged in stasis or when entrained in a flow of fluid, such that imaging data thus obtained can be employed to decode each encoded bead by determining the identity of reporters bound to that bead.

Background of the Invention

Single nucleotide polymorphisms (SNPs) are locations in the genome where a single base substitution has occurred. SNPs are estimated to occur in the human genome at a frequency of approximately 1:1000, implying that there are three million SNPs in the three billion nucleotide human genome. Since most known gene sequences are on the order of 1,000 base pairs in length, each gene is expected to contain one SNP. It is estimated that there are 100,000 human genes, meaning that there are 100,000 SNPs, which may directly affect the function and/or expression of the resulting proteins.

The relative abundance of SNPs in the genome has stimulated efforts to quantify the location and frequency of occurrence of single base substitutions as a tool for the analysis of gene function. Methods for the detection of SNPs include the oligonucleotide ligation assay (OLA), single-strand conformation polymorphism

analysis, allele-specific oligonucleotide (ASO) hybridization, and the single base chain extension (SBCE) assay.

While SNPs are abundant, an individual SNP is not rich in information. Most SNPs are in non-coding or non-regulatory regions of the genome and may not affect gene expression at all. Of those SNPs that occur within the coding regions, many are likely to occur in non-critical regions of the resulting protein or may result in a benign (or nonexistent) amino acid substitution and therefore have little or no bearing on the protein's function. Further, only a fraction of the total number of genes are actively expressed at a given time, so the presence of an SNP within a gene does not indicate *a priori* that it has significant phenotypic relevance. Since the fraction of phenotypically relevant SNPs is small, it is useful to have a high throughput means of analyzing SNPs in order to identify those of biological importance. Once the genome has been thoroughly analyzed and the locations and relative abundances of important SNPs have been identified, a multiplexed method of SNP analysis from an individual's deoxyribonucleic acid (DNA) will be clinically useful. It would therefore be desirable to have a high throughput method for analyzing numerous SNPs in a short period of time.

The locations of SNP and other polymorphic loci within the genome can be determined by sequencing and comparing the same sections of the genome from numerous individuals. Those locations within the genome that are statistically variable across individuals are polymorphic. The biological relevance of a given polymorphism can be determined by correlating the different alleles to the presence of disease or other phenotypic traits. Hence, there is a need for a robust, inexpensive, and widely available method for sequencing gene-sized lengths of DNA in order to discover the locations and biological relevance of polymorphic sites.

One SNP analysis method calls for the binding of oligonucleotides to supports such that numerous identical oligos are bound to a solid support, and so that different supports bear different oligo sequences. One method of encoding an oligonucleotide library useful for SNP analysis is to place unique optical reporters on solid supports during combinatorial chemical synthesis. The attachment of reporters to the supports may be by means of covalent bonds, colloidal forces or other such means, to ensure that reporters stay in contact with, or in close proximity to, the solid support. The solid supports are typically beads of polystyrene, silica, resin, or any another substance on which compounds can be readily synthesized and to which reporters can be affixed in a split/add/pool (SAP) combinatorial process. Each reporter encodes both the identity of a molecular component as well as its place in the synthetic process. By enumerating the optical characteristics of each reporter bound to a solid substrate it is possible to

decode libraries of unique compounds numbering in the billions. As noted above, useful genetic assays can be performed by combinatorially synthesizing oligonucleotides on a bead library such that a given bead bears numerous identical covalently bound oligos and each bead in the library bears a different oligo sequence. In addition to its oligo sequence, each bead bears a unique optical signature comprising a predefined number of unique reporters, where each reporter has a predefined combination of different fluorochromes. A bead's optical signature is correlated to the addition sequence of each reporter during the synthetic process to enable identifying the unique nucleotide sequence on that bead. By imaging the beads, the optical signatures can be read and correlated to the corresponding oligo sequences.

In addition to DNA analyses, the reporter labeling method is also useful for synthesizing diverse libraries of chemical compounds on beads for subsequent analyses in drug candidate screening. Likewise the same method can be used for synthesis of protein libraries on beads where the base unit of synthesis is one of twenty amino acid sequences.

Generally, with only a few reporters and colors, the number of unique signatures that can be created is quite substantial. For example, using only five colors and five reporters, more than 10,000 unique signatures can be generated. Using six colors and 10 reporters, over 115 million unique signatures can be generated to create a very diverse bead library. Clearly, the number of unique combinations that can be identified using reporter labeled beads is substantial. The nature of the apparatus required to identify the unique spectral signatures of such beads is discussed in greater detail below.

Substrate-Based Approach to Analyzing Reporter labeled Beads

In order to read the reporter signature of a bead, an image of the bead must be acquired with sufficient spatial resolution to discriminate the locations of individual reporters. If reporter size or shape are used in the signature scheme, the spatial resolution must be sufficient to discriminate these parameters as well. Further, the acquired image must have sufficient spectral resolution to accurately discriminate the multiple colors emitted from a single reporter. Further still, the quality of the imagery of each bead acquired must be sufficient to ensure that at least one copy of every unique reporter on a bead is evident in the view. Even when multiples of each reporter are bound to a bead, there remains a probability that not every unique reporter will be resolved in a given image. Reporters may not be in clear focus or they may not be exposed to the optical collection system because of their disposition on the bead. In such cases, multiple images should be acquired of each bead at different focal planes or